

DOCUMENT RESUME

ED 157 919

TM 007 258

TITLE Improving Educational Achievement. Report of the National Academy of Education Committee on Testing and Basic Skills To the Assistant Secretary for Education.

INSTITUTION National Academy of Education, Washington, D.C.

PUB DATE [78]

NOTE 26p.

EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.

DESCRIPTORS *Academic Achievement; Achievement Tests; *Basic Skills; Criterion Referenced Tests; Educational Accountability; Educational Assessment; *Educational Trends; Elementary Secondary Education; Federal Government; *Government Role; Norm Referenced Tests; *School Responsibility; State Standards; *Student Improvement; Student Testing

IDENTIFIERS Minimum Competency Testing; Test Score Decline

ABSTRACT

In response to eight questions concerning educational achievement, the National Academy of Education's Committee on Testing and Basic Skills cited four changes in school practices which have caused declines in writing skills and Scholastic Aptitude Test scores. Those changes are a proliferation in courses with less rigorous intellectual standards; confusion about the appropriate methods of instruction; slackening of the amount of learning and teaching classroom time; and a dismantling of opportunities for intensive study in selective academic environments at the secondary level. To enhance educational achievement, the committee recommends a focus on basic skills from ages five to eight, followed by a broader use of skills and curriculum balance in higher grades. Increased emphasis on instructional time or time on task is recommended for performance improvement, as well as challenging material, beginning reading, early diagnosis and remediation, reading to children, mastery learning, computer-assisted and peer instruction. Minimum competency standards are not advised for the high school level, but a series of standardized tests at lower grade levels is recommended for pinpointing student weaknesses and remediation needs. Mandatory and voluntary national tests are discouraged, due to potential interference in local curriculum decisions. Federal research and information dissemination on educational achievement and standardized tests are encouraged. (Author/JAC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Stephen K. Bailey

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

ED157919

improving educational achievement

REPORT OF THE
NATIONAL ACADEMY
OF EDUCATION

TO THE
ASSISTANT SECRETARY
FOR EDUCATION

TM007 258

Foreword

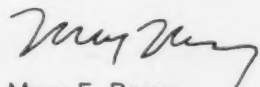
In August of 1977, in response to public interest in the issues of student achievement and testing, I established a Task Force on Basic Skills and Quality Education within the Education Division of the Department of Health, Education, and Welfare. The Task Force considered what additional role the Education Division might play in supporting State and local efforts in basic skills education to improve the usefulness of tests for parents, teachers, and students. Under the leadership of Joel S. Berke, the Task Force served as the forum for deliberations on these issues within HEW. In October, partly as a result of the findings of the Task Force, Secretary Joseph F. Califano, Jr. announced seven initiatives dealing with basic skills and testing; these included my invitation to the National Academy of Education to provide us with advice and recommendations concerning eight specific questions. In my request to the Academy, I indicated that the questions which they were asked to consider were central to our concerns, and that the collective judgments of the distinguished scholars whom they convened would be of immense benefit to us.

Their report amply justifies my expectations. It was a central subject of discussion at the recent National Institute of Education Conference on Achievement Testing and Basic Skills convened on March 1, 2, and 3 in Washington, and attended by more than three hundred leading educators and public officials. In addition, Secretary Califano relied on the document in his explanation of the basis for the proposed program of technical assistance and research that HEW will develop in this area. Finally, the report has been the subject of extensive media coverage across the country.

In my view, the most significant aspects of the Academy Report are its thoughtful analysis of the appropriate and inappropriate uses of tests, its discussion of the promising directions that schools can take to improve learning.

The national concern with basic skills and minimal competency is likely to continue until Americans are persuaded that our schools are providing quality education to all students. The Academy report should help significantly to guide this movement into productive channels. We are, therefore, indebted to its President, Stephen K. Bailey, who served as chairperson of the Panel, and to its members, for preparing this useful document in timely fashion.

The Education Division Task Force included: Mary F. Berry, Chairperson; Joel S. Berke, Deputy Assistant Secretary for Education (Policy Development); Dan Meltzer, Special Assistant to the Secretary; Michael O'Keefe, Deputy Assistant Secretary for Planning and Evaluation (Education); William Blakey, Deputy Assistant Secretary for Legislation (Education); Michael Timpane, Deputy Director, National Institute of Education; Gil Schiffman, Director, USOE Right To Read Program; Frank Corrigan, National Center for Education Statistics; and Richard Hendrix, Fund for the Improvement of Postsecondary Education. Kate Avery and Mary Moore, policy analysts in the Office of the Assistant Secretary for Education, served as staff to the Task Force.



Mary F. Berry
Assistant Secretary for Education
U.S. Department of Health, Education,
and Welfare

Table of Contents

	Page
Preface	i
Executive Summary	iii
Question 1	1
The findings of the Wirtz report are that many factors have been involved in the recent decline in test scores. Have there been, in your judgment, changes in schooling in the past two or three decades that have also contributed to the decline? Such factors as a declining standard, too little instructional time, and changing curriculum have been mentioned. Is there evidence to bolster or to question these and related assertions?	
Question 2	3
Official as well as popular statements are increasingly being made in defense of a return to "basic skills." Do you believe these statements are justified? Are "reading, writing, and numbers" an adequate definition of "basic skills"? If not, what are others? What curricular balance should be sought among the basic skills identified? What are the trade-offs in focusing energy and resources on improving student achievement in the basic skills areas?	
Question 3	5
What are promising directions for schools to take to improve educational attainment levels? Are there differences that are related to income and minority status? What role should standardized tests play? How should they interact with curriculum and school programs?	
Question 4	8
Is the goal of increasing achievement test scores an appropriate, feasible goal for Federal, state and local educational policymakers? If so, what are the anticipated differential impacts by income and SES groups?	
Question 5	9
Present efforts of States to establish "minimum competency standards" inevitably raise questions of appropriate grade levels and cutoff points. What potential problems should be addressed in the States' efforts to institute minimal competency requirements?	
Question 6	11
There is a growing emphasis on holding the schools responsible for what and how well students learn. What is the most likely administrative level—school building, school districts, State, etc.—for producing measurable changes in educational achievement? What role, if any, should the Federal government play in this accountability movement?	
Question 7	12
Is there a role for the U.S. government in producing or sanctioning "voluntary national tests", or in monitoring or assuring the quality of tests used by SEA's and LEA's? If so, what is that role and how should it be structured and with what safeguards?	

Question 8

13

Some scholars challenge the validity of distinctions drawn between "norm-referenced" and "criterion-referenced" tests. What is the significance of this distinction, if any, for public policy?

References

17

Preface

In November 1977 the Secretary of Health, Education, and Welfare, Joseph A. Califano, Jr. and Assistant Secretary for Education Mary F. Berry, reflecting widespread official as well as public concern about declining SAT scores and an alleged decline in the basic-skill competencies of American young people, asked the National Academy of Education to establish a panel on testing and basic skills. The NAEd Panel was composed of Stephen K. Bailey, Graduate School of Education, Harvard University (chairman); John B. Carroll, University of North Carolina, Chapel Hill; Jeanne Chall, Graduate School of Education, Harvard University; Robert Glaser, University of Pittsburgh; John I. Goodlad, University of California, Los Angeles; Diane Ravitch, Teachers College, Columbia University; Lauren Resnick, University of Pittsburgh; Ralph W. Tyler, Science Research Associates, Chicago; and Robert L. Thorndike, Teachers College, Columbia University.

The Academy Panel met with a group of staff experts from HEW on November 16, 1977. Assistant Secretary Berry attended that meeting. Subsequently, Assistant Secretary Berry, on behalf of the Task Force on Basic Skills and Quality Education, submitted to the National Academy of Education Panel a list of eight policy-related questions and asked that they respond within a two-month period. Each member of the Panel reacted in writing to each question. The responses were then collated by Dr. Arthur E. Wise, formerly of NIE. His preliminary draft was circulated to members of the Academy Panel and to selected HEW staff. A colloquium for the draft was slated for February 7, 1978, but had to be postponed because of weather. A smaller colloquium took place on February 16, 1978. The preliminary draft was then reworked into the following report. It represents a distillation of the best thinking of the NAEd Panel. As is inevitable in a complex and evolving field, there emerged differences of opinion in the Panel itself. The body of the report represents areas of agreement. Dr. Kenneth B. Clark, a distinguished social psychologist, a member of the New York State Board of Regents, and a member of the Academy was asked to review and to comment on the report draft. His comments appear as footnotes at the appropriate places in the typescript, as do qualifying comments by Panel members.

The Panel is indebted to Assistant Secretary Berry and Deputy Assistant Secretary Joel S. Berke and his staff for a variety of facilitative and documentary services.



Stephen K. Bailey
Chairman

Executive Summary

Improving Educational Achievement

Report of the National Academy of Education Committee on Testing and Basic Skills to the Assistant Secretary for Education.

1. Changes in Schooling

The Panel believes that four changes in school practices have been particularly important in causing declines in writing skills and SAT scores. First, there has been a proliferation of courses with less rigorous intellectual standards. Second, there has been confusion about the appropriate methods of instruction. Third, there has been a slackening of the amount of classroom time spent on learning and teaching. Fourth, there has been a dismantling of opportunities for intensive study in selective academic environments at the secondary level.

2. "Return to Basic Skills"

American education should be paying much more attention to doing a thorough job in the fundamentals of reading, writing, and arithmetic beginning with ages five to eight, but then expanding the curriculum so that these basic skills are utilized in a broader, more demanding curricular diet. It is at the higher grades that the question of curricular balance becomes important. It is not reasonable to think that secondary schools, for example, will challenge and retain young people on a diet of *only* reading, writing, and arithmetic—even though enriched curricula at higher levels must always provide for the continuous upgrading of basic-skill competencies. And where teachers in the lower grades have not fulfilled their educational responsibilities in the basic skills, teachers at upper-grade levels and in secondary schools must make up for these deficiencies even at the cost of a leaner curricular diet for the learner who requires additional instruction. Otherwise, students will suffer from progressive incapacity to cope in a highly competitive society.

3. Improving Educational Achievement

While there are no panaceas for improving educational achievement, research does suggest some promising directions. There is a striking convergence of evidence that points to the role of time-on-task, "engaged time," etc. in improving performance. Other suggestions include the need for challenging material, a code emphasis in beginning reading, early diagnosis and remediation, reading to children, mastery learning, computer-assisted instruction and peer instruction.

There is no evidence to suggest that anything fundamentally different is needed for low-income or minority children than for any other children.

Different tests are required to serve needs for grouping, diagnosis and accountability.

4. Increasing Achievement Test Scores as a Governmental Goal

The goals of eradicating illiteracy and of increasing educational achievement generally—so long as these do not lead to an over-riding preoccupation with test scores—are surely appropriate goals for the educational system.

There is the risk that the testing of minimal competencies may only be the measuring of accumulated deficits.

5. The Establishment of "Minimum Competency Standards"

The NAEd Panel believes that any setting of state-wide minimum competency standards for awarding the high school diploma—however understandable the public clamor which has produced the current movement and expectation—is basically unworkable, exceeds the present measurement arts of the teaching profession, and will create more social problems than it can conceivably solve.

However, the Panel is in agreement that a series of standardized tests at the lower grade levels used for diagnosing individual student weaknesses, pinpointing remediation needs, and building public pressures if school-wide performances in basic skills continue over time to be consistently low, could be positive influences on student learning. Such tests can have negative effects if they become the sole magnet of educational energies. But used with care, and with increased public (particularly teacher and parental) understanding of the diagnostic and therapeutic use (and limitations) of tests, minimum competency testing can be a positive educational development.

6. Accountability for Educational Achievement

The school district or, preferably, the school should be held responsible for what and how well students learn. It is that administrative level which should be held responsible for producing changes in educational achievement. It is important to hold responsible those who actually have the power to do something. The importance and expense of education tempts the public and politicians to support educational panaceas. Instead, we should be insisting that local schools perform.

The Federal government should support research on the factors which produce measurable changes in educational achievement and should foster the development and dissemination of information designed to increase public (especially parental) understanding of the uses and abuses of standardized tests.

7. Role of the Federal Government

The Academy strongly recommends against the creation of "national tests"—mandatory or voluntary. Federal approval of tests is also risky. The power to approve tests is ultimately the power to approve the curriculum. The logic is clear: those who make up tests that confer educational status on individuals determine what is taught in the local classroom. The Federal government should not be thrust into the position of fostering a national curriculum by inadvertence.

8. Norm-referenced and Criterion-referenced Tests

It would be a mistake to regard "criterion-referenced" tests as necessarily better, more desirable, or more useful than "norm-referenced" tests. It would be wrong to decree that all "norm-referenced" tests be discarded or outlawed, and that only "criterion-referenced" tests be permitted. One and the same test can be both "referenced to criteria" and "referenced to norms." Further, both "norm-referenced" and "criterion-referenced" tests have their uses in education. And, both norm-referencing and criterion-referencing of tests have distinct and relevant functions in education.

Question 1

The findings of the Wirtz report are that many factors have been involved in the recent decline in test scores. Have there been, in your judgment, changes in schooling in the past two or three decades that have also contributed to the decline? Such factors as declining standards, too little instructional time, and changing curriculum have been mentioned. Is there evidence to bolster or to question these and related assertions?

For fourteen years there has been a drop in the average score on the Scholastic Aptitude Test taken by high school students seeking admission to college. There has also been a consistent decline in recent years in the ability of thirteen-year-olds and seventeen-year-olds to write acceptable English.

These declines have understandably triggered both concern and analysis. American education has set itself a most difficult but necessary task: mass literacy of a high order. In analyzing educational successes and failures, educators must concern themselves both with maintaining excellence and assuring those levels of minimum academic achievement that all students should attain. The work of the members of this Panel suggests that greater numbers of people have reached a higher level of literacy than ever before in our history. At the same time, test-score declines suggest a flattening out of the educational pyramid. Thus we see three problem areas which are different but not unrelated: first, those who are left behind by rising standards of minimum literacy; second, declining achievement among the college bound; and third, the overall achievement of American schools. The evidence suggests cause for concern about academic performance of each of these groups.

Whatever non-school factors may have been involved (e.g., TV, parental permissiveness, the size of the age cohort in high school), there is growing evidence that shifts in policy, expectations, and behavior *within schools themselves* have contributed substantially to the documented decline in writing skills and in aptitude test scores. The Wirtz committee, commissioned by the College Entrance Examination Board to study the test score decline, identified in 1977 some salient within-school factors. The NAEd Panel believes that four within-school factors have been particularly important in causing declines in writing skills and SAT scores.

First, the proliferation of courses. For whatever reasons, many of them originating as external pressures upon the schools, the past twenty years have seen a proliferation of subject-matter courses available to pupils in the upper levels of elementary school and in junior and senior high schools. With these increased options, students have been able to choose courses less demanding than the traditional regimen of English, Math, Science, and Social Studies (especially History). The proliferation of non-intellectually rigorous courses has, unfortunately, been accompanied by a decline in the intellectual standards demanded by formerly rigorous, still required courses.

Second, confusion about the appropriate role of teachers. The 1950s and 1960s were periods of mass questioning about the ends and means of education. Teachers were given new and often contradictory models of appropriate pedagogic behavior. "The discovery method," "the open classroom," "individualized instruction," "team teaching," became "in" words, but they were rarely accompanied by teacher training and professional development programs adequate for effective implementation. In effect, teachers were confused. They floundered between what they knew how to do and what they were more and more being expected to do, but not prepared to take on effectively. There was very little good traditional teaching and a paucity of creative approaches to the new.

Third, slackening of "on-task" attention. Harnischfeger and Wiley have learned that time spent in school is directly related to school achievement. But even more important is the percentage of time actually spent by students applying themselves to scholastic tasks.

Some studies suggest that even in relatively "good" classrooms students are "on task" for only about 30% of the instructional day (and the instructional day in a "good" classroom is devoted to learning only 60% or 70% of the time).

Goodlad, Klein and Associates suggest that an extraordinary amount of school time is spent on transitions from class to class, routines, announcements, interruptions, and general control and discipline.

Fourth, a dismantling of opportunities for intensive study in selective academic environments at the secondary level. In city after city programs for the academically talented have been closed or forced to accept changed admission standards. Islands of excellence within our public schools are rarer today than twenty years ago. We believe that this has contributed to the decline in numbers of students performing at the highest levels on SAT and similar tests. We need to improve instruction in the elementary schools to the point where access to certain programs at the secondary school level can be competitive without resulting in economically and ethnically segregated classrooms.

Although comparative studies are hard to come by, the NAEd Panel is convinced that "on task" attention has slipped in the past two decades, and that more effective use of school time in whatever the curriculum calls for would be a most significant reform in American schools.

Question 2

Official as well as popular statements are increasingly being made in defense of a return to "basic skills." Do you believe these statements are justified? Are "reading, writing, and numbers" an adequate definition of "basic skills"? If not, what are others? What curricular balance should be sought among the basic skills identified? What are the tradeoffs in focusing energy and resources on improving student achievement in the basic skills areas?

Two realities underscore the importance of re-emphasizing "basic skills," assuming this phrase includes a learning core in reading, writing, and numbers. First, is the deplorable and declining quality of written composition among all sectors of young people in this country. Second, is the heavy penalty that functional illiteracy exacts from the twenty percent of American youth who come from homes where parents have had little or no education—homes largely concentrated in inner cities and isolated rural areas.

The term "basic skills" at its periphery is subject to a variety of meanings. Most people would agree, however, that reading, writing, and numbers are at the cognitive core. Experts disagree about how developed these skills must be in order to have individuals claim that they are truly competent in them. And many people would argue that certain affective (i.e., socially oriented) skills are essential if children are to become useful citizens and are to be prepared adequately for their various roles in society. The Academy Panel would only underscore the proposition that the distinctive role of the school lies in the development of cognitive functioning. It is the only social institution that has the development of cognition as its primary purpose. The basic skills are the building blocks for the more complex skills of scientific and value analysis; technological and occupational coping; historical, psychological, and philosophical understanding; and political participation. Without the acquisition of the fundamental learning tools of reading, writing, and numbers, no education of value can proceed.** Furthermore, a complex technological and bureaucratic society demands a high level of literacy from all those who would participate usefully and successfully in its functioning. The standard is higher than it ever has been before, and should be; and more people are expected to reach it, and should be. In such circumstances, a renewed emphasis on basic skills at these higher standards seems to this Panel eminently justified.

More specifically, American education should be paying much more attention to doing a thorough job in the fundamentals of reading, writing, and numbers beginning with ages five to eight, but then expanding the curriculum so that these basic skills are utilized in a broader, more demanding curricular diet. It is at the higher grades that the question of curricular balance becomes important. It is not reasonable to think that secondary schools, for example, will challenge and retain young people on a diet of *only* reading, writing, and arithmetic—even though enriched curricula at higher levels must always provide for the continuous upgrading of basic-skill competencies. And where teachers in the lower grades have not fulfilled their educational responsibilities in the basic skills, teachers at upper-grade levels and in secondary schools must make up for these deficiencies even at the cost of a leaner curricular diet for the learner who requires additional instruction. Otherwise, students will suffer from progressive incapacity to cope in a highly competitive society.

**It should be noted that reading and writing implies facility with and growth in the related language skills of speaking and listening.

This is not to say that basic skills are the be-all and end-all of education. Clearly they are not. They are essential but not sufficient. Schools bear a large share of the burden of developing advanced analytic, critical, and artistic skills. Oral skills as well as the basic language skills of reading and writing need cultivation. Schools share with other community institutions a responsibility for developing in young people various social skills. An effective curricular balance to accommodate these various public expectations must be carefully structured at the upper levels of schooling. But such a balance is unthinkable without adequate basic skills acquired in the primary and elementary years.

The issue of "trade-offs" in focusing energy and resources on improving student achievement is specious. If a good foundation has been set in basic skills, the further honing of such skills and the acquiring of an expanded universe of knowledge at higher grade levels are mutually reinforcing—at least for the well-taught student.

Question 3

What are promising directions for schools to take to improve educational attainment levels?

The question seems to assume relatively simple universal answers. Actually there are no panaceas, even if there are promising findings and developments.

Extrapolating from research findings, the answer to the question of how schools can improve educational attainment lies *in spending more time on those attainments we value*.

There is a striking convergence of evidence that points to the role of time-on-task—"engaged time"—in improving performance in school subject matters. This line of evidence, however, needs interpretation, for it is subject to a great deal of oversimplification and to potentially regressive actions on the part of schools eager for quick (and cheap) increases in test scores. At first glance, the evidence seems to call for a return to group, teacher-led instruction, since this is a known way to keep most children at work a good percentage of the time.¹ But the evidence suggesting that more time spent leads to stronger performance is, virtually without exception, evidence based on group means. It shows that, *on the average*, groups with more exposure will do better on tests than groups with less exposure. It does not say that *every child* does better simply by lengthening time on task. Two groups might not profit from this: the very bright, for whom the extra time is redundant (they might profit more from being allowed to move ahead in a program, or at least being left free to read and explore on their own) and the relatively slow, who may not follow what is going on in group instruction aimed at the "middle", and, thus, although present, not be really "on task" in an intellectual sense. These children need to work, too, but on something within grasp. Thus, some degree of individualization seems to be needed.

Learning problems need to be diagnosed in individuals at an early age and related remediation needs to be administered before children fall hopelessly behind. In short, more attention needs to be paid to individual progress, not only to a group mean. In statistical terms, more attention must be paid to the fine points of distributions of scores and not only to means and medians. In the rush to show test score gains, the temptation to attend only to means, and to use the easiest and least expensive methods of raising average scores, may lead to giving up much of what has been painfully learned in the last fifteen years in the way of individual planning and personalized teaching.

There is another potential loss in too quick a return to the old-fashioned group-organized instruction. That is a loss of a new and more open social environment which has permeated many schools as more individual activity has become normal. Perhaps the "open classroom" movement pushed the matter too far for most people, but it is probably also the case that most Americans would regret having to return to the old "line them up in rows" authoritarian classroom in return for increasing test scores.² Perhaps the tradeoff is not required. If time-on-task evidence is analyzed coolly, it argues that *any* organization of schools that keeps time-on-task high will solve the achievement problem. If time-on-task can be increased while simultaneously giving children reasonable degrees of control over their own time and learning, *both* increased achievement and responsive social environments may be possible. Some computer-aided and other programmed instruction have provided at least partial positive clues to resolving this dilemma.

¹ Dr. Ravitch does not feel that getting students to spend more time on task necessarily implies in any way a return to group, teacher-led instruction.

² Dr. Ravitch believes that teachers should control and direct the learning process, but that this need not imply a cold and unresponsive social environment in the school.

Other promising directions include the following:

- With regard to level of challenge, it appears that *difficult programs lead to higher achievement levels than easy programs.*
- *For beginning reading, a "code"³ (learning connections between sounds and letters) emphasis seems to produce better results than a "meaning" emphasis.*
- *Early diagnosis and remediation are more effective than later intervention.*
- *Reading to children, particularly on a level above their own linguistic development, is related to further linguistic development and to reading development.*
- *A competent and energetic school administration is essential for pupil achievement—in setting standards of expectation, in providing reading specialists, in making proven instructional materials available, etc.*
- *In some subjects, the "mastery learning" formats of Benjamin Bloom and others show great promise—although their cost in adequate teacher training and the difficulties of enforcing extra-classroom remedial exercises should be squarely faced.*
- *In some settings, peer instruction has had a profound effect on educational achievement.*

There are, however, no easy substitutes for well-trained and committed teachers, principals who are helpful to and supportive of staff, a disciplined atmosphere in schools, improved pupil nutrition and health, and ultimately the raising of the standards of living of the poor in our midst.

Are there differences that are related to income and minority status?

There is no evidence to suggest that anything fundamentally different is needed for low-income or minority children than for any other children. They, too (perhaps they especially), need to get down to work on academic tasks. Such success stories as we have with minority and low-Socio-Economic-Status groups all point to the power of a businesslike atmosphere in which structured academic work gets done. The problems (overlooking individuals in the search for average gains) and the tradeoffs will probably be much the same for these groups as for any others. It is indeed possible that as we become more refined in our efforts to teach more to everyone, different approaches will turn out to work better for different subgroups in the population. But we are doing so little real teaching now, especially to low-Socio-Economic-Status and minority children, that we can probably show enormous initial improvement by simply attending to the elementary task of doing more teaching and keeping expectations high.⁴ The fine tuning can come later.⁵ There is no surer way to keep minority children from a fair chance than teaching them something different from what everyone else is being taught (or not teaching them much of anything at all). Note that this does not imply that we should pay no attention to cultural differences of children. If children

³ Dr. Goodlad objects to this statement. He does not feel that the evidence is in on this issue. He does not know of even one study in which teachers knew what they were doing that resulted in a decline in reading scores. Teaching for meaning is very difficult, but, when carefully done, the fundamental skills do not suffer. He fears that this is just one more instance of teachers being able to handle decoding but not being able to handle reading for meaning. He believes that precisely the same situation prevails in mathematics. When teachers do a good job of teaching mathematics for meaning, the basic skills do not suffer. He would regret taking a position that may discourage teachers from using more creative methods when the evidence is not yet in.

⁴ Italicized for emphasis at Dr. Clark's suggestion.

⁵ Dr. Ravitch does not agree that "higher expectations" can succeed where other approaches have failed. Instead of depending on "fine tuning", specific programs must be encouraged to aid faltering urban systems.

speak no English, there is no point trying to teach them to read English as if they did. Their special situation must be realistically dealt with—perhaps by longer hours and more weeks in school each year since there is so much more that they have to learn.

What role should standardized tests play? How should they interact with curriculum and school programs?

Tests can play several different roles. One is as a means of public accountability, a way in which everyone interested can keep track of whether the schools are actually doing much teaching. The other is in helping to manage the instructional process, by either grouping students or diagnostically analyzing their competence so that instruction can be matched in relatively fine-grained ways to the needs of the moment. The trouble is that the major standardized achievement tests we now use are not ideally suited to any but the grouping function. For public accountability, they seek to spread out scores arbitrarily and magnify individual (and thereby social class and ethnic) differences—and may therefore mask real gains especially in low groups. In addition, they are expensive to administer. A much shorter test, perhaps using matrix sampling techniques in which not all children take all items, would suffice for accountability purposes, and some progress in the development of such tests now seems to be on the horizon. For diagnostic purposes, again the item selection techniques of our main standardized tests by and large make them unsuitable for diagnosis. The items are not "clean" enough; and there are often too few on any given skill or topic. Furthermore, giving one big test is not a very efficient way of doing diagnosis, other than for initial grouping purposes.

Question 4

Is the goal of increasing achievement test scores an appropriate, feasible goal for Federal, state and local educational policy-makers?

The goal of increasing test scores *per se* is a trivial one, and could possibly become a pernicious one.⁶ That is, if the goal leads to coaching for the test or to efforts to "beat the test" it can pervert education. The possibility cannot be ignored.

The goals of eradicating illiteracy and of increasing educational achievement generally—so long as these do not lead to an overriding preoccupation with test scores—are surely appropriate goals for the educational system.

The question is: should such goals be described as local, state, or national? And what are the policy implications of so construing them? Whose responsibility is it to increase educational achievement? We note that none of the "promising directions" described in question 3 requires Federal or state action to be implemented. Of course, their implementation can be improved and facilitated and encouraged by Federally-funded research and development.

If so, what are the anticipated differential impacts by income and Socio-Economic-Status groups?

There is the risk that the checking of minimal competencies may only be the measuring of accumulated deficits. The consequences upon those who do not pass must be assessed. If success on tests (for purposes of graduation or promotion) is achieved by 4/5 of a suburban school system, system, but only 1/3 of a central-city system, the consequences could be serious for domestic tranquility as well as social equity in a world where a high school diploma, regardless of intrinsic meaning, is frequently a ticket to particular jobs.

If social policy increases the trauma associated with educational sorting devices by increasing the stigmas of scholastic failure, more is at stake for the society at large than the elevation of average standards.

⁶ Dr. Clark believes this first sentence to be unnecessary and irrelevant. Increasing test scores just for its own sake is not the goal of either parents or educators.

Question 5

Present efforts of states to establish "minimum competency standards" inevitably raise questions of appropriate grade levels and cutoff points. What potential problems should be addressed in the states' efforts to institute minimal competency requirements?

The NAEd Panel believes that any setting of state-wide minimum competency standards for awarding the high school diploma—however understandable the public clamor which has produced the current movement and expectation—is basically unworkable, exceeds the present measurement arts of the teaching profession, and will create more social problems than it can conceivably solve.⁷

It is basically unworkable because in many populous states cut-off points for a passing grade that are politically and educationally acceptable to parents, pupils, and educators would have to be so low that an overwhelming majority of students would be allowed to pass. This would make the diploma standard almost meaningless. The recently reported experiences of the Board of Regents in New York State in not approving their own mandated high school graduation test because it was too low, is a classic case of what will happen in one form or another in all states. The present measurement arts of educational testing are simply not up to the ambiguous expectations reflected in most state legislation. In a field like English composition, for example, objectives and testable standards are subject to an almost infinite variety of professional disagreements and technical quibbles.

Setting minimum competency standards for high school graduation is the final step in a completely unsuccessful effort to reduce to precise behavioral terms what education is all about. Attempts to impose competency-based teacher education and competency-based preparation of school administrators have surely failed. The effort to determine and assure minimum competency standards for high school graduation will also fall of its own weight, for the scaffolding of existing test designs is too weak to carry such an emotionally-laden and ambiguous burden. Continuing any extensive efforts and funds in this direction is wasteful and takes attention away from the major tasks of improving our schools.

As noted elsewhere in this report, the social consequences of failing those in our society who have accumulated educational deficits does not cure *their* problems. It only exacerbates them.

However, *the Panel is in agreement that a series of standardized tests at the lower grade levels used for diagnosing individual student weaknesses, pinpointing remediation needs, and building public pressures if school-wide performances in basic skills continue over time to be consistently low, could be positive influences on student learning.*⁸ Such tests can have negative effects if they become the sole magnet of educational energies. But used with care, and with increased public (particularly teacher and parental) understanding of the diagnostic and therapeutic use (and limitations) of tests, minimum competency testing can be a positive educational development.

⁷ Dr. Clark disagrees with the thrust of this response. He argues that minimum competency testing may only measure the "accumulated deficits", but one has to know what the deficits are in order to correct them. How will we know if the findings and directions that are so well stated on page 6 are being accomplished if we don't test? Mere attendance and physical "time-on-task" are not enough. We need some indication that children are learning everything they need to learn, and that those held responsible for instruction are really teaching what they should. Dr. Clark agrees that minimum competency testing for high school graduation *alone* is not the answer and that early testing when remediation is still possible is certainly preferable. However, he feels that we should test at the high school level rather than not test at all in order to assure at least a minimum standard of education for everyone.

⁸ Italicized for emphasis at the suggestion of Dr. Ravitch.

"Going public" with state-wide minimum competency tests at lower and medium grade levels is useless unless it stimulates better teaching. Unfortunately, there is not today a situation in which the public seems willing to trust the judgments of teachers and principals in certifying their children's competence. To that extent we have as a nation "bought" the argument of the original test builders that judgments of competence would be more reliable and less biased when based on scientifically constructed tests than when left to the judgments of school people. It is not only the test-builders who believe this. School people themselves probably believe it, and many parents and other lay people are adamant about it.⁹ So tests are needed. But what kinds? The term "minimum competency" implies a criterion-referenced rather than a norm-referenced test. But the validity—and social acceptability—of such a test depends upon acceptance of the definitions of competence that are built into the tests. How can agreement on these matters be secured?

There seem to be two contradictory routes being followed. One is to use the global standardized tests that now exist. Because they are global there is general agreement that they are testing the right kinds of things. They may work for skills such as reading or computation. They will not work so well as soon as knowledge rather than skill is at stake (for example, in history or science), because there will be many different ideas of what content should be taught. Tests that sample widely across content might be a compromise solution, socially workable, although not very "clean" or attractive to instructional purists.

The second approach has been to ignore the global tests and start building lists of competencies that people think are important for children to acquire. Often these lists are developed by or in conjunction with citizens' committees. The politics of the idea is superficially attractive, but the result is a *reductio ad absurdum* of the basic idea of criterion-referenced standard setting. Knowledge and skill cannot be divided into hundreds of discrete objectives; valid tests for all those separate objectives can probably not be written; it would take impossible amounts of time to test everyone on all those objectives; and we would not know what to do with the information once we got it.

We are then forced back to the concept of general tests for minimum competency in a very limited number of basic skills. *As has been pointed out, using such tests for high school graduation alone is bound to fail. Experimenting with critical "competency points" at selected lower grade levels has substantial merit.*¹⁰

⁹ Dr. Ravitch does not agree that the public necessarily disdains educators and distrusts their judgments. Some confidence in a teacher's ability to make basic judgments about a student must be maintained if we are to rebuild the effectiveness and legitimacy of educational institutions. Testing need not be the *only* answer.

¹⁰ Italicized for emphasis at the suggestion of Dr. Ravitch.

Question 6

There is a growing emphasis on holding the schools responsible for what and how well students learn. What is the most likely administrative level—school building, school districts, state, etc.—for producing measurable changes in educational achievement?

The school district or, preferably, the school should be held responsible for what and how well students learn. It is that administrative level which should be held responsible for producing changes in educational achievement. It is important to hold responsible those who actually have the power to do something. The importance and expense of education tempts the public and politicians to support educational panaceas. Instead, we should be insisting that local schools perform.

What role, if any, should the Federal government play in this accountability movement?

The Federal government should support research on the factors which produce measurable changes in educational achievement and should foster the development and dissemination of information designed to increase public (especially parental) understanding of the uses and abuses of standardized tests.

Question 7

Is there a role for the U.S. government in producing or sanctioning "voluntary national tests", or in monitoring or assuring the quality of tests used by state education agencies and local education agencies? If so, what is that role and how should it be structured and with what safeguards?

There is a broad array of possible Federal roles with respect to testing. At one extreme is a single national mandatory test for high school graduation. At the other extreme is providing information. There may be one test or many. Test use may be mandatory or voluntary. Test development may be publicly or privately financed. Public financing may be at the Federal, state, or local level. Test development may be under public or private auspices. Quality control may be publicly or privately financed and managed. Test administration may be national, state, or local; financing of test administration may be Federal, state, or local. Moreover, the results of the test may be used for making decisions about individual students, classrooms, schools, school districts, states, or the nation. The results may also not be used for making decisions.

A mandatory national test would have the Federal government interfering with the responsibility of the states for education. Why should the Federal government produce a national test? Is it to indicate what the Federal government thinks the schools should be teaching? If the tests are to provide information about the progress of American education, the National Assessment of Educational Progress accomplishes that purpose without infringing on the states' responsibility.

The notion that the Federal government should be the Bureau of Standards for educational testing is both professionally unsound and politically dangerous. Since tests reflect the educational goals for many students and teachers, the Federal approval or disapproval of tests is in effect having the Federal government determine what the schools should teach. It was this fear that led the Executive Committee of the American Association of School Administrators to oppose the National Assessment.

While acknowledging the legitimacy of the Federal government as a partner in improving educational opportunity and quality, the Academy wishes to underscore the importance of maintaining the principal Constitutional responsibility of the states in this important area of public concern.

*In short, the Academy strongly recommends against the creation of "national tests"—mandatory or voluntary.*¹¹ Federal approval of tests is also risky. The power to approve tests is ultimately the power to approve the curriculum. The logic is clear: those who make up tests that confer educational status on individuals determine what is taught in the local classroom. The Federal government should not be thrust into the position of fostering a national curriculum by inadvertence.

Federal funding of agencies which develop tests or of agencies which engage in quality control of tests is less problematical. The key is that a *variety* of such centers be supported. What is important is to avoid a single, official, Federal testing monolith.

And, of course, the Federal government can and should support research designed to improve the quality of tests, and programs to increase public understanding about their use.

¹¹ Dr. Clark thinks that the sanctity of state and local control is overemphasized. He contends that our population is too mobile for inferior education to remain contained within a given state. We must have some minimum *national* standard of educational attainment—it must be at a level higher than each individual state. A rejection of national standards leaves the path open for no standards at all. The danger of no standards should outweigh the fear of too much Federal control.

Question 8

Some scholars challenge the validity of distinctions drawn between "norm-referenced" and "criterion-referenced" tests. What is the significance of this distinction, if any, for public policy?

Challenging the Distinctions between the Tests

Those who challenge the validity of the distinctions between "norm-referenced" and "criterion-referenced" tests argue that norm-referencing and criterion-referencing are *processes* that can be applied to tests. Either of these processes can be applied to one and the same test. A test is "criterion-referenced" or "norm-referenced" depending upon which of these processes has been applied to it.

Norm-referencing simply means that statistical operations have been performed on test data to make possible the interpretation of the scores in terms of "norms", that is, in terms of the standing or rank of any given score in a distribution of scores that has been determined for some given sample or population of students to whom the test might appropriately be given. For example, a particular score on a reading test might be found to stand very high among all scores made by third-graders in urban schools—say, at the 95 percentile, meaning that it stands above 95% of all scores made by students in that population, but below the top 5% of all scores of those students. The same test score might be found to stand much lower in some other group—say, at the 60th percentile for a representative sample of *fourth*-graders in urban schools, or for a sample of third-graders in *suburban* schools.

Over many years, norm-referencing has been found to be a convenient and generally well-understood way of interpreting test scores. It permits assessing a student's educational progress in comparison to the level or degree of progress that students of a given age or grade typically make. Using test norms, a teacher can readily confirm that a child is making markedly slower progress than other children of the class. A school principal can compare the progress of third-graders in his or her school with that of third-graders across the country. In many circumstances, such comparisons are valid and useful in judging whether efforts to teach or educate students are attaining expectations.

Norm-referencing of tests does have certain drawbacks, however. It depends upon the availability of suitable normative data. Good normative data are difficult to collect; publishers of standardized tests often spend large amounts of money in the effort to establish representative and up-to-date norms, but norms can easily become out-of-date or inaccurate for a number of reasons. Norm-referencing inevitably implies comparison of a student with other students. While these comparisons may be fair when the student is compared to others in the student's own norm group, they may not be completely fair, and they may even become invidious, when students are compared to students in norm groups other than an appropriate one. For example, it might be regarded as unfair to compare urban students' reading scores with those of students in suburban schools. (On the other hand, such a comparison might be thought by some people to be very useful.) The major drawback of norm-referencing, however, is that it does not give concrete information on exactly what level of ability, knowledge, skill, or performance is represented by a given test score, even if that test score is a highly accurate measure. Thus, while it may be informative to know that a third-grade student attains the 60th percentile among third-graders on a standardized reading test, this does not tell how well the student can read; i.e., what level of text difficulty and complexity he or she can handle. This kind of information would be forthcoming, if at all, only if the test is also criterion-referenced.

Criterion-referencing means that the scores on a test can be interpreted in terms of concrete, specifiable, and describable levels of ability, achievement, or performance. That is, a criterion-referenced test score is one that can be interpreted as indicating that the student possesses some specified level of knowledge, understanding, or skill. Sometimes this means going into minute details as to exactly what the student knows or does not know—or can or cannot do. Scores on an arithmetic test, for example, might be "criterion-referenced" in the sense that they would indicate exactly what addition facts the student has failed to master. On the other hand, a much more generalized idea of what a test score means is frequently desired. It would be too much to expect a reading test score to indicate exactly what words a student could recognize; there are simply too many words to be tested, and even if one could develop a test that could be "criterion-referenced" in this way, the information might be almost overwhelming for a teacher. Rather, it would be desirable to "criterion-reference" a reading test in terms of the representative levels of text or passage difficulty that the student getting a given score could be expected to comprehend. If one can decide what level of text difficulty represents "functional literacy", the test could be said to yield information on whether a given student has attained functional literacy.

From the above, one can see not only that it is possible either to "norm-reference" a test or to "criterion-reference" it, but also that it is possible for a test to be *both* "norm-referenced" and "criterion-referenced", depending upon what kinds of research operations have been performed in connection with the test.

In principle, norm-referencing is comparatively simple—even though it may require large and costly research operations. To "norm-reference" a test, it is administered to large and representative samples of one or more defined groups that the test is likely to be given to, or is appropriate for—e.g., third-graders, fourth-graders, and fifth-graders, or to urban third-graders, suburban third-graders, rural third-graders, etc. Score distributions are then set out in each of the groups so that the test user can find the level of a given score with reference to an appropriate group. These score distributions are usually presented in "tables of norms", and there are various systems offered for interpreting scores—percentiles, "standard scores", "T-scores", and the like.

Criterion-referencing a test is in many ways much more difficult than norm-referencing it. It requires a detailed analysis of what kinds and amounts of knowledge are represented by a given score level on the test. Obviously, a very high score would indicate that the student has mastered all or nearly all the knowledges or skills tapped by the test, and a very low score would indicate that the student has failed to master all or nearly all of those knowledges and skills. Even so, interpreting these scores entails communicating exactly what knowledges and skills are or are not measured by the test, and this is often not easy to do. Most good tests are designed on the basis of a detailed specification of the abilities, knowledges, or skills they are intended to measure, but this does not always guarantee that those abilities, knowledges, or skills are in fact measured by the test. Even with painstaking care on the part of the test maker, scores sometimes reflect abilities and achievements (or lack thereof) that were not intended to be measured. For example, a test designed to assess knowledge of the grammar of a foreign language might turn out to be much more a measure of knowledge of the vocabulary of that language.

Even though it might be relatively simple to interpret very high and very low scores on a test, it is harder to "criterion-reference" scores in the middle ranges of the distribution. Sometimes it is possible to arrange the skills and knowledges tapped by a test in order of relative difficulty, and then to establish a more or less close correspondence between the test scores and the ranking, in terms of relative difficulty, of those skills and knowledges. A medium-level score, then, might be found to mean that the student getting such a score has mastered the easier skills, but has failed to master the more difficult skills. Thus, a medium-level score on a spelling test might mean that

the student can spell correctly all or nearly all of a list of easily spelled words, but would fail to spell correctly many words that are often misspelled. Note that in this case criterion-referencing would entail giving detailed lists of words that students at various score levels could or could not spell.

This kind of criterion-referencing is generally unfeasible for the traditional types of standardized achievement tests in the various school subjects—reading, arithmetic, science, social studies, etc. For one thing, these tests attempt only to include small samples of all the skills and knowledges that are ordinarily taught in these subjects. It would be difficult to put scores on such tests in correspondence to highly-detailed lists of specific skills and knowledges that students at different score levels would possess. And even if this were done, the results would be so voluminous that it might be burdensome for teachers to utilize them effectively. Only when criterion-referencing of achievement tests can use rather generalized descriptions of skill and knowledge levels is it possible to "criterion-reference" most of the standardized tests commonly used in today's schools. Such criterion-referencing has not generally been attempted, but it is quite feasible, and certainly desirable. For example, a standardized test of competence in a foreign language can be criterion-referenced in terms of generalized proficiency; e.g., "able to read simple prose", "able to carry on a conversation on a general topic with a native speaker of the language", etc. Similar generalized descriptions could be given for ranges of score levels on achievement tests in other school subjects.

Because of what are perceived as difficulties in criterion-referencing of traditional types of tests in the regular school subjects, some educational measurement specialists have recommended the development of tests that would lend themselves readily to criterion-referencing. Generally, this has meant the development of a large number of very brief tests—even individual items or "exercises"—that are specifically targeted on measurement of particular educational objectives. For example, one might have a test to determine whether a student has mastered a particular skill (e.g., the operation of "carrying" in adding a column of numbers) or understands a particular concept (e.g., "the division of labor"). In this sense, and only in this sense, one can say that there are certain types of tests that are specifically "criterion-referenced" and not "norm-referenced". Yet, one could also, as it were, "norm-reference" these tests by finding out what percentages of various groups can pass them. (A good illustration of a type of "norm-referencing" of these highly targeted tests is afforded by the sample results published by the National Assessment of Educational Progress on what percentages of different groups, by age, grade, and region of the country, can pass particular exercises included in the NAEP assessments.)

Highly-targeted "criterion-referenced" tests of specific skills and achievements are more useful in the day-to-day conduct of teaching than they are in the overall evaluation of students' progress at the end of the school year or at the end of a major segment of the student's educational career. Standardized general achievement tests are usually more suitable for the latter purpose.

Supporting the Distinctions between the Tests

Those who argue that norm-referenced tests are different from criterion-referenced tests stress that distinctions between the tests arise from the selection of behavior to be sampled, the definition of the universe of the test. This defined universe is the "criterion." If the criterion is clearly defined and the universe properly sampled a trained user of the test can judge more rationally its usefulness for the purpose he has in mind than is possible with a norm-referenced test. As an example, the lack of definition of the universe sampled by the SAT made it impossible for the Review Panel to judge the significance of the score decline in any terms other than self-contained statistical ones.

Implications for Public Policy

It would be a mistake to regard "criterion-referenced" tests as necessarily better, more desirable, or more useful than "norm-referenced" tests. It would be wrong to decree that all "norm-referenced" tests be discarded or outlawed, and that only "criterion-referenced" tests be permitted. One and the same test can be both "referenced to criteria" and "referenced to norms." Further, both "norm-referenced" and "criterion-referenced" tests have their uses in education. And, both norm-referencing and criterion-referencing of tests have distinct and relevant functions in education.

References

- Anderson, R.C., & Adams, M.J. An Analysis of Cognitive Processes in Early Reading. *Journal of Language Arts*, in press.
- Bennett, N. *Teaching Styles and Pupil Progress*. London: Open Books, 1976.
- Berliner, D.C., & Rosenshine, B. The Acquisition of Knowledge in the Classroom. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the Acquisition of Knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Bond, G.L., & Dykstra, R. The Cooperative Research Program in First-Grade Reading. *Reading Research Quarterly*, 1967, 2 (entire issue).
- Chall, J.S. *Learning To Read: The Great Debate*. New York: McGraw-Hill, 1967.
- Chall, J. with Conrad, S. and Harris, S. *An Analysis of Textbooks in Relation to Declining SAT Scores*. Prepared for Advisory Panel on Score Decline. New York: College Entrance Examination Board, 1977.
- Chomsky, Carol. "Stages in Language Development and Reading Exposure," *Harvard Educational Review*, 42 (1), February 1972.
- Denckla, M. "Minimal Brain Dysfunction," in Chall, J.S. and Mirsky, A.F. (Eds.), *Education and the Brain: The 77th Yearbook of the National Society for the Study of Education*. Chicago: NSSE, 1978.
- Fischer, C.W., Filby, N.N. & Marliave, R.S. *Instructional Time and Student Achievement in Second Grade Reading and Mathematics*. Paper presented at the meeting of the American Educational Research Association, New York, April 1977.
- Glaser, R. *Adaptive Education: Individual Diversity and Learning*. New York: Holt, Rinehart & Winston, 1977.
- Goodlad, John I. "A Perspective on Accountability." *Phi Delta Kappan*, 57 No. 2, October 1975, 108-112.
- Goodlad, John I. "What Goes on in Our Schools?" *Educational Researcher*, 5 No. 3, March 1977, 3-6.
- Goodlad, John I. "What Schools Are For." A Transatlantic Dialogue. The Sir John Adams Lecture, University of London Institute of Education, October 18, 1977.
- Goodlad, John I. and Klein, M. Frances. *Looking Behind the Classroom Door*. Belmont, Ca.: C.A. Jones, 1974.
- Guthrie, J.T., Samuels, S.J., Martuza, V., Seifert, M., Tyler, S.J., & Edwall, G. *A Study of the Locus and Nature of Reading Problems in the Elementary School*. (Final Report, 2 sections). Newark, Delaware: International Reading Association, 1976.
- Harnischfeger, A. and Wiley, D.E. *Achievement Test Score Decline: Do We Need To Worry?* Chicago: CEMEREL, Inc., 1975.
- Just, M., & Carpenter, P. (Eds.). *Cognitive Processes in Comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.
- Kohl, H. *Reading, How To*. New York: Bantam Books, 1973.
- Newsletter, October 1976—National Assessment of Educational Progress, Denver, 1976.
- Olson, D.R. From utterance to text: "The Bias of Language in Speech and Writing." *Harvard Educational Review*, 1977, 47 (3), 257-281.
- On Further Examination*. Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline, College Entrance Examination Board, New York, 1977.
- Popp, H.M. and Lieberman, M. "A Study of the Relationship of Student Achievement to Components of Reading Programs and Environmental Characteristics." Final Report #400-75-0064, National Institute of Education, 1977.
- Resnick, D.P., & Resnick, L.B. "The Nature of Literacy: An Historical Exploration." *Harvard Educational Review*, 1977, 47 (3), 370-385.
- Rosenshine, B. "Classroom Instruction." In N.L. Gage (Ed.), *The Psychology of Teaching Methods: The 75th Yearbook of the National Society for the Study of Education*. Chicago: NSSE, 1976.

- Rozin, P., & Gleitman, L.R. "The Structure and Acquisition of Reading." In A.S. Reber & D. Scarborough (Eds.), *Reading: Theory and Practice*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.
- Stallings, J. *Implementation and Child Effects of Teaching Practices in Follow-Through Classrooms*. Monographs of the Society for Research in Child Development, Vol. 40, nos. 7-8, December 1975.
- Stebbins, L.B., St. Pierre, R.S., Proper, E.C., Anderson, R.B., & Cerva, T.R. *Education as Experimentation: A Planned Variation Model* (Vol. IV-A: An Evaluation of Follow-Through). Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1977.
- Tyler, Ralph W. "Are Our Children Learning Less?" Chicago, Center for the Study of Democratic Institutions, August 26, 1977.
- Tyler, Ralph W. "The Total Educational Environment." *Report of the National Task Force on Citizenship Education*. New York: McGraw-Hill, 1977.